DOCUMENT RESUME

ED 073 142 TM 002 388

AUTHOR Penfield, Douglas A.
TITLE Educational Statistics.
INSTITUTION Educational Testing Service, Princeton, N.J.; ERIC Clearinghouse on Tests, Measurement, and Evaluation, Princeton, N.J.
SPONS AGENCY National Inst. of Education (DHEW), Washington, D.C.
REPORT NO ERIC-TM-18
PUB DATE Feb 73
NOTE 11p.; 1972 AERA Conference Summaries

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS Conference Reports; *Data Analysis; *Educational Research; *Mathematical Models; Speeches; *Statistical Analysis; *Statistics

ABSTRACT

The 30 papers in the area of educational statistics that were presented at the 1972 AERA Conference are reviewed. The papers are categorized into five broad areas of interest: (1) theory of univariate analysis, (2) nonparametric methods, (3) regression-prediction theory, (4) multivariable methods, and (5) factor analysis. A list of the papers reviewed, their authors, and, when applicable, the ED numbers concludes the summary. (DB)

ED 073142

ERIC CLEARINGHOUSE ON TESTS, MEASUREMENT, & EVALUATION
EDUCATIONAL TESTING SERVICE, PRINCETON, NEW JERSEY 08540

Conducted by Educational Testing Service in Association with Rutgers University Graduate School of Education

TM Report 18

February 1973

1972 AERA Conference Summaries

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY

TM 002 388

EDUCATIONAL STATISTICS

Douglas A. Penfield
Rutgers University

This publication was prepared pursuant to a contract with the National Institute of Education, U. S. Department of Health, Education and Welfare. Contractors undertaking such projects under government sponsorship are encouraged to express freely their judgment in professional and technical matters. Points of view or opinions do not, therefore, represent official National Institute of Education position or policy.

PREVIOUS TITLES IN THIS SERIES

1. Developing Criterion-Referenced Tests
   ED 041 052

2. Test Bias: A Bibliography
   ED 051 312

3. Ability Grouping: Status, Impact, and Alternatives
   ED 052 260

4. Developing Performance Tests for Classroom Evaluation
   ED 052 259

5. Tests of Basic Learning for Adults: An Annotated Bibliography
   ED 058 274

6. State Educational Assessment Programs: An Overview
   ED 058 309

7. Criterion Referenced Measurement: A Bibliography
   ED 060 041

8. Measures Pertaining to Health Education: I. Smoking
   ED 060 042

9. Measures Pertaining to Health Education: II. Drugs. An Annotated Bibliography
   TM 002 078 (ED Number not yet available)

10. Measures Pertaining to Health Education: III. Alcohol. An Annotated Bibliography
    TM 002 079 (ED Number not yet available)

11. 1971 AERA Conference Summary
    I. Evaluation: State of the Art
    ED 060 043

12. 1971 AERA Conference Summary
    II. Criterion Referenced Measurement
    ED 060 134

13. 1971 AERA Conference Summary
    III. Educational Statistics
    ED 060 133

14. 1971 AERA Conference Summary
    IV. Test Development, Interpretation, and Use
    ED 060 135

15. 1971 AERA Conference Summary
    V. Innovations in Measurement
    ED 060 044

INTRODUCTION

About 700 of the 1,000 papers presented at the 1972 AERA Annual Meeting in Chicago, Illinois were collected by the ERIC Clearinghouse on Tests, Measurement, and Evaluation (ERIC/TM). ERIC/TM indexed and abstracted for announcement in *Research in Education (RIE)* 200 papers which fell within our area of interest - testing, measurement, and evaluation. The remaining papers were distributed to the other Clearinghouses in the ERIC system for processing.

Because of an interest in thematic summaries of AERA papers on the part of a large segment of ERIC/TM users, we decided to invite a group of authors to assist us in producing such a series based on the materials processed for RIE. Four topics were chosen for the series: Criterion Referenced Measurement, Evaluation, Statistics, and Test Construction.

Most papers referred to in this summary may be obtained in either hard copy or microfiche form from:

ERIC Document Reproduction Service (EDRS)
P.O. Drawer 0
Bethesda, Maryland 20014

Prices and ordering information for these documents may be found in any current issue of *Research in Education.*

EDUCATIONAL STATISTICS

Douglas A. Penfield

Once again the AERA annual meeting has produced an abundance of research papers in the area of educational statistics with content covering a wide range of theory and application. For purposes of discussion, the educational statistics papers which were submitted to ERIC are broken into five broad areas of interest: (a) theory of univariate analysis, (b) nonparametric methods, (c) regression-prediction theory, (d) multivariable methods, and (e) factor analysis. As was true at the 1971 AERA convention, research on factor analytic methods represented the most frequently discussed topic.

THEORY OF UNIVARIATE ANALYSIS

Papers in this section range from testing assumptions under a univariate model to a discussion of a randomized block design for dichotomous variables. Not too technical in nature, they can easily be read and understood by researchers having a minimum of mathematical training.

Ramseyer and Tcheng discuss the robustness of the studentized range statistic, q, with request to violation of the assumptions of normality and homogeneity of variance. Using an IBM 360/50 computer, values of q were generated for groups (k) of 3 and 5, each containing 5 and 15 scores respectively. When k = 3, the homogeneity of variance assumption was violated by allowing one of the sample variances to become two and then four times as large as the variance in the other two groups. For k = 5, the variance in two of the samples was allowed to become two and four times larger than the variance in the three remaining samples. The normality assumption was altered by transforming scores into distributions which were positively and negatively skewed, exponential, and rectangular. Comparisons are then made between Type 1 error rates at the .05 and .01 levels.

The results show that when the homogeneity of variance assumption is violated, the observed Type 1 error rates are slightly higher than the established rate. Under violation of the normality assumption, the observed error rates are generally below the fixed level. For simultaneous violation of normality and equal variance, the error rates are larger than the nominal levels selected. Nevertheless, the overall variation between observed and expected error rates is minimal and closely resembles results obtained when the assumptions are satisfied.

The study is particularly interesting because comparisons on the distribution of q at the .05 and .01 levels are also made when all the assumptions are met. This adds credence to the random sampling procedure which was developed for the study.

A two group–two treatment experimental design is presented by Maxey. The notation and layout are similar to the format developed by Campbell and Stanley (1963). The series of events for each group consisted of two initial observations at different points in time, followed by a first treatment, an observation, a second treatment, and a final observation. The advantages and purposes of the design are discussed in detail.

Since the effects are confounded over treatments, the author proposes to set up a 2-way repeat measures design to estimate needed error variances. Once the error variances are determined, he uses them to evaluate planned comparisons of interest developed around various combinations of cell means.

An example is presented to illustrate a practical use for the design, but it has a number of undesirable features, including a small sample size (N = 21), and an inappropriate use of the error terms for setting up a number of planned comparisons. Thus, it is impossible to place much faith in the F values so computed. The method of analysis proposed for this extremely complex design cannot be justified mathematically.

Byars and Roscoe describe a procedure for transforming uniformly distributed data into data having an approximate normal distribution. The authors point out that this procedure would find its greatest value in Monte Carlo type studies where uniformly distributed pseudo-random numbers are generated by a computer and must subsequently be normally transformed.

Using the standard normal cumulative distribution function, P (z), an algebraic approximation to the inverse Gaussian is derived. The authors consider this procedure to be more accurate and computationally efficient than the algebraic approximations developed by Hastings and Burr. The Byars-Roscoe approximation involves the use of rational polynomial expressions under a linear data transformation. When P (z) ranged between .01 and .99, it was found to have greater accuracy and require less time to compute than any of the other previously discussed procedures.

The implications of pooling the interaction term with the error term in a 2-way factorial design when testing for main effect differences are discussed by Pohlmann. After reviewing the pros and cons of pooling, Pohlmann describes a Monte Carlo study to illustrate the consequences of this procedure on tests of main effects. The variation of parameters described in the Monte Carlo study are somewhat difficult to follow, but the effects of the pooled error term on main effect differences are observed over changes in a non-centrality parameter, sample size and alpha level.

The results indicate that when the ratio of the interaction degrees of freedom to the error degrees of freedom is less than 0.08 and the interaction term is not significant at the .25 level, it may be useful to pool the interaction term with the conventional error component. In so doing, one creates a test of main effect which is considered to be slightly more powerful than the test which uses the standard error term. Due to a lack of breadth in the development of the Monte Carlo study, the results are not generalizable beyond a 2-way factorial design.

Draper investigated the problem of employing analysis of variance procedures to analyze dichotomous repeated measures data. Two situations are presented; one in which dichotomous responses are gathered on the same items at four separate occasions and a second where the responses occur on different items over the four occasions. The situations can be represented as a 3-way factorial design with items isolated as a contributing source of variation.

All sources of variation, as well as possible confounding effects and appropriate F-tests, are discussed in detail. Using simulated data, a Monte Carlo study was set up with variations made in the base probability of a one, number of subjects and the degree of heterogeneity of the subjects. Comparisons were then made between normally distributed data and dichotomous data. The results indicate that the power of the analysis of variance test based upon dichotomous data is less than one half of the power of the same test performed on normal data. Draper suggests that if the dependent variable is dichotomous, one should choose a large sample to insure reasonable power, for he obtained the greatest power when the probability of a one was near 0.5 and there were 6 or more subjects in the experiment.

In a study similar to the Draper one, Mandeville makes a comparison among three methods of analyzing dichotomous data under a randomized block design. Studies summarizing Cochran's Q test and comparing it with the F test are noted. Mandeville, using dichotomous data, then makes comparisons between F, Q, and a multivariate test statistic, M, developed around Hotelling's T^2.

Dichotomous data were simulated from a multivariate normal distribution and comparisons were made between the empirical and theoretical distributions of F, Q and M. In general, for varying numbers of treatments and blocks, the F statistic has a smaller average error than either Q or M, with M being the least desirable of the three. With respect to power, F is consistently superior to Q. The M statistic is not used to make power comparisons. On the basis of the findings, the F test is recommended over Q or M when: (1) the total sample size is greater than 60, (2) the interrelationship between variables is constant, and (3) the data is believed to come from an underlying normal distribution.

Draper's and Mandeville's studies are both well conceptualized and executed, although the one by Draper is slightly more global in nature.

NONPARAMETRIC METHODS

Recently, research into nonparametric statistical methods has received a minimum of attention from behavioral scientists. One explanation may be the large number of studies investigating and confirming the robustness of many of the parametric procedures currently in use. Of the papers received for review by ERIC, only two could be placed in the category of nonparametric methods. One deals with describing and comparing some nonparametric tests useful for testing equality of variance in the two sample problem, and the other investigates some Chi-square and Kolmogorov models for testing goodness of fit to normal.

Penfield contracts three nonparametric tests for scale, focusing on their method of development and usage. He chose the Siegel-Tukey test, the Mood test and the Normal Scores test, but paid special attention to the Normal Scores test because of its excellent power relative to the parametric F test. Procedures for computing the test statistic for large and small samples are outlined in detail for all three tests of dispersion. Two examples, one for small N and the other for large N, are presented. Results are then computed and compared for the three tests. In the case of the large sample example, the normal approximation to the exact test is illustrated, and power comparisons are made. Of the tests considered, the power of the Normal Scores test is greatest when scores are drawn from distributions having sharp tails. For distributions having heavy tails, the Siegel-Tukey test is superior to the other nonparametric tests.

A study of the robustness of the Chi-square and Kolmogorov statistics under the linear score scale and equal areas models is reported by Kittleson and Roscoe.

The authors restrict themselves to investigating the goodness of fit of data relative to a normal distribution. They present a brief review of the literature on the use of the Chi-square test for goodness of fit and the Kolmogorov test.

Using normally distributed and uniformly distributed random numbers generated by a computer, Chi-square and Kolmogorov statistics are computed on samples under the linear score scale and equal areas models for varying numbers of sample sizes and cells. When comparing nominal and empirical Type 1 error rates, the Chi-square equal areas model proves superior to all other tests. The Kolmogorov tests are found to be very conservative and considered inferior to the Chi-square tests. The best power is obtained when the number of cells approximates 20. This study is clear, concise and well-executed.

REGRESSION–PREDICTION THEORY

Of the papers reviewed in this category, the most prevelant topic pertained to the development of linear programming models. There was also considerable interest in the use of regression analysis to answer questions normally investigated by means of analysis of variance and analysis of covariance.

Referencing Cohen's work on contrast coding for multiple linear regression models, Lewis and Mouw extend the work to include orthogonal comparisons. The authors have broken their discussion into two parts, first showing how analysis of variance models can be written in regression form, and then treating analysis of covariance models in a similar fashion. In the case of analysis of variance, discussion is restricted to one-way and two-way designs, and orthogonal coefficients for setting up trend contrasts under the regression model are introduced. An illustration of different arrays of coding coefficients in the predictor vectors is presented for various contrasts of interest. The authors recommend this procedure over the conventional analysis of variance because the use of independent predictor vectors accurately reflects the degrees of freedom for the analysis, and also permits the investigation of specific contrasts of interest, in addition to the depiction of overall main effect differences.

In a similar fashion, one-way and two-way models under the analysis of covariance are described. The authors give specific attention to pooling the interaction with the error term when the interaction is not found to be significant. They note that contrast coding does not require this pooling, thereby yielding identical results to the traditional analysis of covariance model.

Greenberg and Mejias investigate a use of linear least-square multiple regression analysis with dummy variables for isolating the effect of the individual teacher on student achievement. The sample under investigation consists of 572 students enrolled in a social science course at Miami Dade Junior College. Independent variables used for prediction purposes consist of an English Aptitude and a Social Science score on the Florida State-Wide Twelfth Grade Test (F.T.G.), grade point average, class size, cumulative hours earned, and dummy variables representing instructor input. The dependent variable is the student's final exam score in the social science course. Dummy variables are used to determine whether differentiated instruction accounts for variation in the student's final exam scores.

Results indicate that the Social Science score on F.T.G., grade point average, and cumulative hours earned account for the greatest sources of variation in final exam scores, explaining 48% of the total variance. A significant difference is found between instructors and, on the basis of the results, it was possible to rank order instructor performance. Class size was not related to final exam score. Furthermore, neither salary nor salary-related indices were significantly correlated with teacher's contribution to student achievement. Limitations involved with the design and analysis of the study are thoroughly discussed.

AID-4, automatic interaction detector, is described by Koplyay as a procedure for identifying optimal configurations of predictor variables for criterion prediction under a restricted multiple regression model. Instead of starting with a full multiple regression model, AID-4 starts with the group as a unit and through a splitting process maximizes the between sum of squares for variable categories while minimizing the error sum of squares. The value of this procedure lies in its ability to maximize the proportion of explained variance in the criterion variable without having to identify all the interaction components that are present under the full model. For regression analyses built around a large number of predictor variables, AID-4 identifies the generally small subset of these variables which proves to be significant. A branching process developed around the outcomes derived from this program makes it possible to give a more meaningful interpretation to the results. As a help to potential users, an explanation of some of the more useful and informative features of the AID-4 output is also presented. This procedure is certainly noteworthy and should be of major interest to those researchers who prefer to let a machine aid them in the decision making process.

Schnittjer attempts to develop a linear programming model which would be useful for prediction and then tests its accuracy by comparing it with the standard

curvilinear multiple regression equation. Weights were assigned to the levels of each variable so that the difference between actual and predicted scores would be minimal. The objective function was the sum of these differences and was subject to constraints described as person, variable and level width. An example is presented using 66 subjects and is based upon 10 independent variables and 4 dependent variables, each analyzed separately. Independent variables were divided into levels ranging from 5 to 27. Following comparisons between the linear programming and curvilinear multiple regression models, subsets of 50 individuals were randomly selected and their linear programming results compared with results from the 16 individuals not included in the sample. The author concludes that the two models give comparable results. The study is extremely vague and provides no hint as to the nature of the actual equations computed. A test of accuracy is not indicated, which suggests that it was made by the "eyeball" method.

A linear programming model designed to make optimal assignment of students to attendance centers is presented by Ontjes. An object function is developed which minimizes the distance students must be bused in order to reach their assigned centers. Some constraints on the system are the capacity of the school building, grade capacity, and the need to assign everyone within an area to one school. Model and constraint equations are laid out in detail. An example illustrating the use of the model on junior high school students is presented, the purpose of which is to minimize busing while providing a good racial balance. The example is interesting because of its implications relative to current demands being placed upon school systems. Using results generated by a computer, a summary shows the average distance travelled and minority percentage within each school. The study is clear, easy to read and has some definite practical application.

Matzke formulates a linear programming model to simulate a foundation-type support program. The model is then applied to a state support program for the public schools in Iowa. Five objective functions were developed in order to minimize several derivatives of the state mandated local tax rate, to minimize state aid costs of the foundation's program, and to maximize the foundation's level of support. The general linear programming model is stated mathematically, as are the constraints on the model. These constraints fell into 3 general categories entitled district, system, and variable interaction. Inputs to the linear programming model consisted of data obtained from each school district and the Iowa State Department of Education. The equations are analyzed by a computer and the results for each optimization problem are reported. The data analysis is quite extensive and involves solutions for the distribution of funds to a foundation-type program. Tables which show the results under the optimal solution are also provided. The bibliography at the end of the paper is excellent.

Under some linear models the values of the independent variables are assumed to be fixed rather than random. Calkins and Jennings investigate the effects of violating this assumption for a simple linear regression model by observing the number of incorrect decisions made when testing slope and intercept differences under the assumption that the concomitant variable is random instead of fixed. Values of the concomitant variable were drawn from both a normal and rectangular distribution and results were computed on simulated samples of varying size. Sampling distributions for slopes and intercepts were generated and then compared using critical statistics.

The results are elaborately laid out in tabular form. It suffices to say that violation of the fixed variable assumption does not produce significant observed and expected differences with respect to intercepts and slopes. To insure robustness, sample sizes greater than thirteen should be chosen.

Friedman examines the postulate that improved prediction of multiple criteria can be achieved through the use of pattern analysis rather than conventional regression models. Pattern analysis in this instance implies a restructuring of the data so as to increase the accuracy of prediction. This restructuring is handled primarily through the use of factor analysis.

The author hypothesizes that: (1) for a single criterion, simple linear combinations of predictor variables will perform as well as a combination of linear and nonlinear variables; and (2) for predicting multiple criteria, nonlinear and linear combinations of predictor variables will yield a greater multiple correlation coefficient than a simple linear combination of the independent variables.

Data was collected on 700 subjects over 10 scales of the Parent-Child Relations Questionnaire and 6 scales of the California Achievement Test. Analysis consists of finding factor scores, canonical correlations and multiple correlation coefficients with respect to males, females and the total group. Results indicate that a simple linear combination of variables gives the best predication of a single criterion. When predicting multiple criteria, the results are not as clearly defined since nonlinear variables appear to add valuable information to the prediction process.

MULTIVARIATE METHODS

This section is the most diversified of the five areas being covered. It contains papers on such complex topics as time series analysis, multivariate analysis of variance, discriminant analysis, and interaction analysis. The texts range from sophisticated mathematical notation to simple descriptions of empirical research. Because of their informative nature, a number of these papers would be valuable reading in an advanced educational statistics course.

Raw gain scores, residual gain scores and adjusted scores derived from an analysis of covariance are compared by Williams, Maresh and Peebles for the two sample problems. These comparisons are empirical in nature and are based on reading scores obtained by 165 pupils attending rural North Dakota schools. Using notations outlined under a full and restricted multiple regression model, the authors formulate the F statistic for the analysis of covariance. Besides gain scores and the analysis of covariance, the authors recommend the use of residual gain scores for determining differences between the two groups. Their proposed value lies in the fact that they are uncorrelated, can be defined precisely, and lend themselves to determination of higher ordered residual gains. The three procedures were used to test for group differences on reading related variables across grades 2 through 6. Aside from finding differences in the outcomes derived from the three methods of analysis, little significant knowledge is gained from the comparisons being made.

Sachdeva proposes a multivariate analog of Hays' omega squared for estimating the strength of relationship in a multivariate analysis of variance. The term represents the proportion of variation in dependent variable scores which is accounted for by the independent variables used in the study. Omega squared for the univariate case is transformed to the multivariate situation by replacing sums of squares with the determinant of the corresponding matrix of sums of squares and sums of cross-products. Multivariate omega squared is then shown to be a function of Wilk's lambda test criterion. The author also shows that it can be written as a function of an F ratio. An example is presented which explains how to solve the various formulas derived for multivariate omega squared. The procedures are clearly outlined and make a valuable contribution to multivariate methods.

The prediction of teacher turnover using time series analysis was researched by Costa. Two-year and three-year moving averages, as well as exponential smoothing using a 0.1 and 0.9 smoothing factor, were the assessment techniques which he employed. Demographic data was combined with time-series forecasting methods for prediction purposes. The demographic variables of sex, age, marital status, and years of experience were used to identify different types of teachers. Moving average and exponential smoothing results were obtained on each teacher type, and the percent accuracy in predicting turnover rate is reported. Generally the accuracy was above 40 per cent. Not unexpectedly, young married women with less than 4 years of experience showed the highest rate of turnover. The author's use of the Chi-square test of independence to test the equality of non-independent proportions over the four methods is inappropriate, and information as to which technique would give the best prediction was inconclusive.

Rogers investigates the utility of the jackknife for establishing confidence intervals on and testing hypotheses about the disattenuated correlation coefficient for small samples. If a person's score is conceived in terms of two components designated as true score and error score, a disattenuated correlation expresses a relationship between the true scores on two different instruments. Following an extremely thorough review of the literature, the jackknife procedure is explained and illustrated. Essentially it is a procedure for obtaining approximate confidence intervals when standard statistical procedures cannot be applied.

Using a computer to simulate data, sampling distributions of disattenuated correlation coefficients were obtained for different combinations of input parameters. Characteristics of these distributions with respect to central tendency, variability, skewness, and kurtosis are described in detail. The best results were obtained from sampling distributions which were approximately normally distributed and had a large variance. For developing confidence intervals when N is small, the jackknife was found to be superior to procedures based upon normal theory.

Huberty and Blommers compare three indices of predictor variable potency in order to ascertain the contribution of each variable toward the discrimination process over repeated sampling. The indices were: (1) the scaled weights of the first Fisher-type discriminant function; and, (2) the total and within groups correlation estimates between each predictor variable and the first Fisher-type function. After describing the formulation of multiple group discriminant analysis procedures, the criterion for assessing stability of predictor variable potency is discussed in detail. Essentially it involved the observation of variable rank consistency over repeated replications of the experiment. Relationships among the ranks were determined by computing Kendall's coefficient of concordance. The indices based on correlation estimates were found to be somewhat more reliable than the one computed on scaled weights. The rankings were so scattered however, that unless the sample size was very large, none of the indices of variable potency could be relied upon to give consistent results.

A discussion of interaction analysis and how it relates to a one-dependent Markoff chain is presented by Pena. The purpose of investigating the relationship between these two procedures is to test the order of dependence of the interaction chain and to evaluate empirically the power of Darwin's criterion, as well as show its relevance to educational situations. Darwin's Likelihood Ration Criterion tests whether two or more matrices composed of conditional probabilities are equal.

Data collected on sixth grade teachers over five subject areas is used to test for a one-dependent chain among events. Using a Chi-square test of significance it was found that a two-dependent model provides a better fit to interaction data than the one-dependent model based upon Darwin's criterion. A development of the Likelihood Ratio statistic for a Markoff chain of order two is also presented, and possible adjustments in Darwin's criteria in order to reliably analyze data on a one-dependent chain are discussed. The sensitivity of the criteria was evaluated by observing the power of the test. For the conditions imposed, power was found to be very near one. A comparison between the empirical distribution of Darwin's criteria and the Chi-square distribution reveals a close fit for a sequence length of 500.

FACTOR ANALYSIS

The papers in this section represent a wide variety of current research under the broad heading of "factor analysis." Principal component analysis is utilized and discussed frequently, especially in the work of Hakstian. Topics researched using factor analytic methods include an assessment of students' ratings of courses and instructors, the study of relationships between cognitive abilities tests and concept attainment measures, and the comparison of different measures of association.

Through the use of principal component analysis, Magoon and Price researched the factor dimensions produced from student ratings of course and instructor characteristics. They hypothesized that the rated characteristics reflect the raters' preconceptions of course and instructor interrelationships and are not necessarily related to actual course characteristics and instructor behavior.

A very thorough review of the literature is followed by an analysis of three sets of rating data obtained from an instrument consisting of 22 items. One set of data consists of between class ratings, another, within class ratings, and a third set was completed prior to the start of the course. All sets are submitted to a principal component analysis, with the first four unrotated factor dimensions being compared by means of Tucker's coefficient for factor congruence. Interrater reliability is computed for selected raters and was somewhat low. Results show the principal component loadings to be quite similar across samples. The authors conclude that the ratings reveal more about student preconceptions than the framework of meaningful instructional quality.

Plans for investigating the relationship between some cognitive abilities tests and concept attainment measures are reported by Harris. The eventual purpose is to identify those cognitive abilities that are related to concept attainment in four subject matter areas. Three approaches to analyzing the relationship between two sets of variables are outlined: (1) lumping all variables together and factor analyzing; (2) factor analyzing one fundamental data set and correlating variables in the other set with the factor scores; and (3) employing canonical variate analysis and interbattery factor analysis to check the stability of factors over different test selections. A summary of projected computations is presented at the conclusion.

Keown and Hakstian compare five measures of association with respect to stability and robustness of correlation and rotated factor matrices for seven point Likert scale data. The five measures of association are: (1) Pearson's r; (2) tetrachoric r; (3) phi coefficient; (4) phi divided by phi max statistic; and (5) Kendall's Tau-B. Data conforming to 20 Likert scale variables for five different distributions is generated by the computer. The distributions chosen for study are referred to as normal, rectangular, central, positive skew and mixed skew. For each one, five correlation matrices corresponding to the five measures of association are generated among the Likert scaled variables. All correlation matrices are then subjected to a principal component analysis and rotated factor matrices are obtained.

Three measures of robustness are computed from the correlation and component pattern matrices. Results from the normal distribution are compared with findings from the four other distorted distributions. Comparisons between the distributions for the two procedures indicate that Tau-B followed by Pearson's r are least affected by distribution distortion. The effects on each measure of association are discussed in detail. Those measures which initially require a splitting of the data at the median not only create a loss of information, but also produce correlational and factorial results which are less than optimal.

To evaluate the degree of goodness of fit of patterns derived from a principal component analysis, Skakum, Maguire and Hakstian develop an empirical sampling distribution of the average trace statistic and use the statistic to look for similarities between component

structures. They discuss several approaches to factor congruence and look at the differences that exist between two matrices following rotation.

From a population component score matrix, pairwise samples of 50 component scores were randomly selected. The average trace was then computed for each pairwise sample. The sampling distribution of the average trace was found to be positively skewed; consequently, a square root transformation was applied to each trace to create an approximate normal distribution. Properties of the transformed data are reported, and three examples are presented to illustrate application of the average trace statistic.

Hakstian develops a number of factor analytic strategies for handling longitudinal data collected on the same individuals over two different occasions. The five models introduced vary with respect to the stability of component scores and factor pattern matrices. Elaborate descriptions and derivations of each model are presented; four are developed by least square methods, while a fifth relies upon canonical correlation procedures. Empirical examples are computed for situations where: (1) component scores and factor pattern matrices are constant, (2) component scores are constant and factor pattern matrices are variable; and (3) component scores and factor pattern matrices are variable. In the first two examples data was simulated on a computer. Correlations are computed within and between occasions, between true and estimated component scores, and between true and estimated rotated pattern matrices. Results indicate a high degree of correspondence between true and estimated scores.

Researching further into factor theory, Hakstian and Muller discuss ways of determining the significant number of factors in a behavioral experiment. They start by summarizing the explanatory and taxonomic views of factor analysis. A review of three standard factor analysis models defined as component, incomplete component, and common-factor is then followed by a compilation of literature pertaining to the number of potential factors over n variables. Major work by Guttman and Kaiser is given special attention.

As a prelude to the experimental study, the authors review and compare some of the more commonly used rules for determining the appropriate number of factors. Data from seventeen correlation matrices appearing in the literature are analyzed using eight different rules for finding the number of factors. Finding little agreement between the various procedures, the authors conclude that the appropriate number of factors depends upon the factor analytic model and procedures selected, as well as the interpretability of the factors.

Related to the paper by Hakstian and Muller is one by Dzuiban and Harris in which they empirically evaluate the meaningfulness of components in a principal component analysis. They point out that extracting components where the eigenvalues are greater than one may not always produce interpretable results. Bartlett's Test of Sphericity is recommended as one safeguard against performing an inappropriate principal component analysis, but it too is fallible. Lack of statistical significance with respect to Bartlett's test implies that principal component analysis may be an inappropriate method for analyzing data.

Citing data previously analyzed by principal component analysis after Bartlett's test was found to be significant, the authors indicate that a meaningful interpretation of two of the components is in question. They recommend selecting another model and proceed to reanalyze the data using image component analysis, uniqueness rescaling factor analysis and alpha factor analysis. In this illustration, image analysis is found to offer protection against interpreting random variables as forming the basis of a meaningful component. The paper is noteworthy and should be a warning to all potential users of principal component analysis.

Pruzek, Stegman and Pfeiffer discuss a general method for analyzing data which has been partitioned in clusters. They study the relationship between partitions in order to evaluate structural similarities. A measure of the goodness of fit of an empirical cluster of items to some theoretical cluster is developed, and properties of the proposed test statistic, q, are discussed. To illustrate the method, they use an example in which 50 students are asked to partition 26 items into at least five and not more than nine categories. Two different target partitions are selected for purposes of analysis: target 1 is an a priori splitting of items on the basis of specific item characteristics, whereas target 2 is based on results from a latent partition analysis. Values of the test statistic tend to be smaller for target 2. The authors also outline strategies for studying partitions relative to other methods outlined in the literature.

In summary, 30 papers were reviewed under the heading of educational statistics. For purposes of convenience in reading, they were broken down into five broad areas of interest. The two areas receiving the most attention were factor analysis and regression theory. The quality and rigor of the research presented in 1972 appeared to be superior to presentations of a year ago. Perhaps we are making progress after all.

PAPERS REVIEWED

Byars, J.A., & Roscoe, J.T. Rational approximations of the inverse Gaussian function. 7p. (ED 064 395, MF and HC available from EDRS.)

Calkins, D.S., & Jennings, E. An empirical investigation of some effects of the violation of the assumption that the covariable in analysis of covariance is a mathematical variable. 29p. (ED 064 359, MF and HC available from EDRS.)

Costa, C.H. The prediction of teacher turnover employing time series analysis. 28p. (ED 063 336, MF and HC available from EDRS.)

Draper, J.F. A Monte Carlo investigaton of the analysis of variance applied to non-independent Bernoulli variates. 40p. (ED 064 343, MF and HC available from EDRS.)

Dziuban, C.D., & Harris, C.W. On the extraction of components and the applicability of the factor model. 12p. (ED 063 334, MF and HC available from EDRS.)

Friedman, D. The use of pattern analysis for the prediction of achievement criteria. 58p. (ED 062 366, MF and HC available from EDRS.)

Greenberg, B., & Mejias, R. A study of the use of multiple regression with dummy variables to identify instructor contribution to student achievement. 29p. (ED 064 360, MF and HC available from EDRS.)

Hakstian, A.R., & Muller, V.J. Some empirical findings concerning the number of factors problem. 31p. (ED 061 270, MF and HC available from EDRS.)

Hakstian, A.R. Some notes on the factor analytic treatment of measures obtained on two different occasions. 30p. (ED 061 252, MF and HC available from EDRS.)

Harris, C.W. Interbattery factor analysis and canonical correlation analysis as tools for relating concept attainment measures and cognitive abilities measures. 13p. (ED 064 348, MF and HC available from EDRS.)

Huberty, C.J., & Blommers, P.J. An empirical comparison of three indices of variable contribution in multiple group discriminant analysis. 19p. (ED 064 363, MF and HC available from EDRS.)

Keown, L.L., & Hakstian, A.R. Some notes on the choice of measure of association for the component analysis of Likert scale data. 22p. (Document not yet available from EDRS.)

Kittleson, H.M., & Roscoe, J.T. An empirical comparison of four Chi-square and Kolmogorov models for testing goodness of fit to normal. 8p. (ED 064 318, MF and HC available from EDRS.)

Koplyay, J.B. Multiple regression analysis and automatic interaction detection. 9p. (ED 064 362, MF and HC available from EDRS.)

Lewis, E.L., & Mouw, J.T. The use of contrast coding to simplify ANOVA and ANCOVA procedures in multiple linear regression. 17p. (ED 061 276, MF and HC available from EDRS.)

Magoon, A.J., & Price, J.R. Rating dimensions of course and instructor characteristics: The eye of the beholder. 14p. (Document not yet available from EDRS.)

Mandeville, G.K. A comparison of three methods of analyzing dichotomous data in a randomized block design. 38p. (ED 064 347, MF and HC available from EDRS.)

Matzke, O.R. A linear programming model to optimize various objective functions of a foundation type state support program. 45p. (ED 064 331, MF and HC available from EDRS.)

Maxey, J.H. A two group–two treatment research design. 16p. (ED 064 381, MF and HC available from EDRS.)

Ontjes, R.L. A linear programming model for assigning students to attendance centers. 16p. (ED 062 395, MF and HC available from EDRS.)

Pena, D.M. The assumption of a Markoff chain model for interaction analysis. 42p. (ED 064 319, MF and HC available from EDRS.)

Penfield, D.A. A comparison of some nonparametric tests for scale. 16p. (ED 065 559, MF and HC available from EDRS.)

Pohlmann, J.T. The effects of pooling the interaction and within components on the Alpha and power for main effects tests. 12p. (ED 061 282, MF and HC available from EDRS.)

Pruzek, R.M., & Others. On the analysis of partitioned data. 17p. (ED 064 402, MF and HC available from EDRS.)

Ramseyer, G.C., & Tcheng, T.K. The robustness of the studentized range statistic to violations of the normality and homogeneity of variance assumptions. 14p. (ED 062 388, MF and HC available from EDRS.)

Rogers, W.T. Jackknifing disattenuated correlations. 52p. (ED 064 393, MF available from EDRS.)

Sachdeva, D. Multivariate analog of Hays ω^2. 5p. (ED 065 552, MF and HC available from EDRS.)

Schnittjer, C.J. The use of linear programming for prediction. 9p. (ED 061 283, MF and HC available from EDRS.)

Skakun, E.N., & Others. An application of inferential statistics to the factorial invariance problem. 26p. (ED 064 392, MF and HC available from EDRS.)

Williams, J.D., & Others. A comparison of raw gain scores, residual gain scores, and the analysis of covariance with two modes of teaching reading. 17p. (ED 061 287, MF and HC available from EDRS.)